

ADDITIVE ENTROPIES OF PARTITIONS

ADAM PASZKIEWICZ AND TOMASZ SOBIESZEK

ABSTRACT. We provide, under proper assumptions, a description of **additive partition entropies**. They are real functions I on the set of finite partitions that are additive on stochastically independent partitions in a given probability space.

Second version, 2012-02-22. This version looks closer into the notion of continuity. All changes, with the exception of small typographic ones, to the previous version are shown in this colour.

1. INTRODUCTION

The classical discrete theory of entropy concerns itself with real functions H defined on the family of sequences (p_1, \dots, p_n) such that $p_i \geq 0$ and $\sum p_i = 1$. There, the most significant role of all entropies is played by the **Shannon entropy**, given by $H(p_1, \dots, p_n) = p_1 \log \frac{1}{p_1} + \dots + p_n \log \frac{1}{p_n}$. (In here, as throughout the paper, we confine ourselves to base 2 logarithms, as dictated by information theory tradition). It is the only symmetric (i.e. independent of the order of p_i -s) continuous function of such sequences that is normalised by $H(1/2, 1/2) = 1$ and satisfies the following **grouping axiom**¹

$$\begin{aligned} H(a_1 p_1, \dots, a_k p_1, b_1 p_2, \dots, b_l p_2, \dots, c_1 p_n, \dots, c_m p_n) = \\ H(p_1, p_2, \dots, p_n) \\ + p_1 H(a_1, \dots, a_k) + p_2 H(b_1, \dots, b_l) + \dots + p_n H(c_1, \dots, c_m). \end{aligned}$$

Arguably, the grouping axiom can be singled out as the most important and almost defining property of Shannon entropy. Even so, a noticeable part of Information Theory has been played by either some modifications or weakened statements of the grouping axiom. Among them there is the

1991 *Mathematics Subject Classification.* Primary 94A17, Secondary 60A10.

Key words and phrases. additive entropy, inset entropy, partition entropy, axioms of entropy, independent σ -algebras.

This paper is partially supported by Grant nr N N201 605840.

¹This result which is a little better than Shannon's own set of axioms (see [23]) is a version of Fadeev's axioms of entropy, c.f. [14]. The grouping axiom is often called **strong-additivity**. In terms of conditional entropy of random variables it can be succinctly expressed as the so-called **chain rule** $H(X, Y) = H(X) + H(Y|X)$, (see [11], Theorem 2.2.1). The shape of the grouping axiom, can lead us to think about the differences of entropies between two partitions, one coarser than the other. Indeed, a simple, but a very neat reformulation of the grouping axiom in these terms as a certain linearity property can be found in [6]

important and much weaker property, called **additivity**, satisfied not just by Shannon entropy but also by the earlier concept called Hartley entropy (see Example 4 in Section 2, c.f. [16]) and Rényi entropy (see Example 3 in Section 2, c.f. [21]):

$$(1) \quad H(p \otimes q) = H(p) + H(q),$$

where

$$p = (p_1, \dots, p_n), q = (q_1, \dots, q_m), \\ p \otimes q = (p_1 q_1, \dots, p_1 q_m, p_2 q_1, \dots, p_2 q_m, \dots, p_n q_1, \dots, p_n q_m).$$

Every symmetric H that satisfies this equation is called an **additive entropy**. Hartley and Rényi entropies proved themselves useful and applicable in many fields.² Additivity can be found in some axiomatisations of Shannon entropy, one notable instance that also gives an axiomatisation of Hartley entropy is given in [4].

There are quite a few other entropies of sequences and their properties. For their detailed exposition, see [2]. For a modern survey of characterisations of Shannon entropy (among other things), see [12].

Now, the natural setting in which entropies appear in most applications is not on a family of sequences but on a space of events. Indeed, a finite partition can be regarded as a representation of the information carried by experimentally collected data (see e.g. [5]). It seems therefore only fitting and indeed often necessary to introduce the concepts of entropy that take events into account. This has lead to the development of 'mixed theory of information' by Aczél, Daróczy and others. This theory is a research of various so-called **inset entropies** and was introduced in a series of papers, beginning with [3].

Consider a ring \mathcal{B} of subsets of Ω . Before, we were considering a family of sequences of positive numbers (p_1, \dots, p_n) that sum up to 1. Riding roughshod over certain details,³ we now consider a family \mathcal{G} of sequences of pairs (A_i, p_i) such that A_i -s are elements of \mathcal{B} which make up a partition, and p_i -s, as previously, are positive and add up to 1. On such a family we consider functions $I : \mathcal{G} \rightarrow \mathbb{R}$.

Various conditions analogous to symmetry, strong-additivity, additivity, and so on lead on to different inset entropies I . For instance we can define **additive inset entropy** by demanding a version of symmetry and a version of additivity

$$I \left(\begin{array}{c} A_i \cap B_j \\ p_i q_j \end{array} \right) = I \left(\begin{array}{c} A_i \\ p_i \end{array} \right) + I \left(\begin{array}{c} B_j \\ q_j \end{array} \right)$$

²Rényi entropy, for instance, is used in random search [22], coding theory [9], cryptography [8], and differential geometric aspects of statistics [10].

³There are actually several different flavours of inset entropy. For instance, it is sometimes assumed that the ring \mathcal{B} is an algebra (contains Ω) and we consider partitions of Ω , at other times we consider partitions of all sets in \mathcal{B} . We can consider positive p_i -s or just nonnegative. The sets can be nonempty or not necessarily.

Inset entropies have recently found application in considerations involving utility function in gambling, see [18].

We propose a related, yet different, approach. We are given a nonatomic probability space (Ω, Σ, P) . Let \mathfrak{A} be the family of all finite subalgebras of σ -algebra Σ , or in other words algebras generated by finite partitions $\Omega = A_1 \cup \dots \cup A_n$, $A_i \cap A_j = \emptyset$. Consider the following version of additivity

$$(2) \quad I(\sigma(\mathcal{A} \cup \mathcal{B})) = I(\mathcal{A}) + I(\mathcal{B}),$$

for any independent finite algebras \mathcal{A} and \mathcal{B} .

It is the aim of the current paper to examine the family of functions $I: \mathfrak{A} \rightarrow \mathbb{R}$ satisfying condition (2), which we shall call **additive partition entropies**.⁴ It turns out that this research leads to a simple and effective description.

Of course, every additive entropy H can be viewed as an additive partition entropy H_P , $(H_P)(\sigma(A_1, \dots, A_n)) = H(P(A_1), \dots, P(A_n))$ that depends only on probabilities of atoms.

But there are other additive partition entropies. In fact, given $x \in \Omega$ consider the function $L_x: \mathfrak{A} \rightarrow \mathbb{R}$,

$$L_x(\mathcal{A}) = \log \frac{1}{P(A_x)}, \quad \text{where } A_x \text{ is the atom of } \mathcal{A} \text{ that contains } x$$

Naturally, this function satisfies equation (2).

Here is the main result of the paper: “Let I be a continuous additive partition entropy. There exist a continuous entropy H , and a countably-additive set function \mathbf{m} absolutely continuous with respect to probability P such that I is a sum of two continuous additive partition entropies

$$I = H_P + \int_{\Omega} L_x(\cdot) \mathbf{m}(dx), \quad \text{that is}$$

$$I(\mathcal{A}) = H(P(A_1), \dots, P(A_n)) + \sum_{i=1}^n \mathbf{m}(A_i) \log \frac{1}{P(A_i)}$$

Moreover we can assume that $\mathbf{m}(\Omega) = 0$, in this case such a decomposition is unique.”

Let us mention that a somewhat related result, however with much stronger assumptions, and in the setting of additive inset entropy has already appeared in [13], see also [17].

The above result shall be presented as a corollary in Section 5, Theorem 2. Sections 2–4 contain a proof of a slightly more general Theorem 1; namely it is natural to replace the probability P with a somewhat more general notion of a finitely-additive measure P . A follow up paper [25] concerns a similar result in which there are no continuity assumptions.

The proof of Theorem 1 is longer than might be expected. The major difficulty lies in the construction of component $\int_{\Omega} L_x(\cdot) \mathbf{m}(dx)$, i.e. in finding

⁴The term ‘partition entropy’ has been introduced earlier in the context of partitions of a finite set, see e.g. [24]. Additive entropy of partitions has been considered in [7].

the measure \mathbf{m} . The construction, is made of two parts — one is probabilistic in that it depends heavily on the notion of independence (point I below), the other involves algebraic manipulation of measures (points II and III below).

Theorem 1 is a corollary to the following results:

I (cf. Propositions 4 and 5, Section 3) If I is a continuous (see definition 4 below) additive partition entropy then for any events V, W with $P(V) = P(W)$ we can define the number $\Delta(V, W)$ in such a way that the following conditions are satisfied

(a) Whenever there is a partition (A_1, \dots, A_n) with $V \subset A_1, W \subset A_2, P(A_2)/P(A_1) = \lambda$ we have

$$\Delta(V, W) \log \lambda = I(A_1 \triangle (V \cup W), A_2 \triangle (V \cup W), \dots, A_n) - I(A_1, A_2, \dots, A_n)$$

(b) If $P(U) = P(V) = P(W)$ then

$$\Delta(U, W) = \Delta(U, V) + \Delta(V, W).$$

(c) For $V = V_1 \cup V_2, W = W_1 \cup W_2, V_1 \cap V_2 = \emptyset, W_1 \cap W_2 = \emptyset, P(V_i) = P(W_i), i = 1, 2$ we have

$$\Delta(V, W) = \Delta(V_1, W_1) + \Delta(V_2, W_2).$$

(d) $\Delta(\cdot, \cdot)$ is continuous in metric $d(V, W) = P(V \triangle W)$.

II (cf. Lemma 7, Section 3) If $\Delta(V, W)$ defined for $P(V) = P(W)$ satisfies (b)–(d) then there is a unique finitely-additive measure \mathbf{m} , absolutely continuous with respect to P such that $\mathbf{m}(\Omega) = 0$ and

$$\Delta(V, W) = \mathbf{m}(W) - \mathbf{m}(V).$$

III (cf. proof of Theorem 1, Section 4) Assume that for some continuous additive partition entropy I , and measure $\mathbf{m} \ll P$ we have

$$[\mathbf{m}(W) - \mathbf{m}(V)] \log \lambda = I(A_1 \triangle (V \cup W), A_2 \triangle (V \cup W), \dots, A_n) - I(A_1, A_2, \dots, A_n),$$

for $P(V) = P(W)$, and a partition (A_1, \dots, A_n) with $V \subset A_1, W \subset A_2, P(A_2)/P(A_1) = \lambda$. It turns out that $\tilde{I}(\mathcal{A}) := I(\mathcal{A}) - \sum_{i=1}^n \mathbf{m}(A_i) \log \frac{1}{P(A_i)}(\mathcal{A})$ is a continuous additive partition entropy such that $\tilde{I}(\mathcal{A}) = \tilde{I}(\mathcal{B})$ for any \mathcal{A} and \mathcal{B} generated by A_1, \dots, A_n and B_1, \dots, B_n such that $P(A_i) = P(B_i), 1 \leq i \leq n$.

All propositions say something about either continuous or general additive partition entropies and they all follow easily from either continuous-case Theorem 1 or an analogical general result in [25]. The lemmas, however, are of an ‘independent’ nature and some of them, especially Lemmas 7–9, might find themselves useful somewhere else.

2. BASIC NOTIONS AND NOTATION

We shall denote the characteristic function of a set A by 1_A . We shall write $A = \sum A_i$ or $A = A_1 + \dots + A_n$, whenever we have $A = \bigcup_{1 \leq i \leq n} A_i$ and the sets A_i are pairwise disjoint.

Fix a space (Ω, \mathcal{F}, P) with a finitely-additive probability measure P defined on some algebra \mathcal{F} of subsets of Ω . We shall consider the family $\mathfrak{A} = \mathfrak{A}(\mathcal{F})$ of all finite subalgebras of the algebra \mathcal{F} . From now on, we shall hold on to the following assumption regarding measure P

Darboux Property. *For any set $A \in \mathcal{F}$ and any $0 < \theta < P(A)$ there is $B \in \mathcal{F}$ such that $B \subset A$ and $P(B) = \theta$.*

Let us remark that in case (Ω, \mathcal{F}, P) is a usual probability space, the Darboux property is satisfied if and only if the space is nonatomic.

Now, going back to the finitely-additive case, we have a naturally-defined notion of an integral of an $\mathbb{R} \cup \{+\infty\}$ -valued simple function. We shall assume that $0 \cdot (+\infty) = 0$.

For any $K \in \mathcal{F}$ such that $P(K) > 0$, we shall consider a truncated conditional probability space $(K, \mathcal{F}_K, P|_K)$. This means that $A \in \mathcal{F}_K$ is equivalent to $A \in \mathcal{F} \wedge A \subset K$ and that

$$P|_K(A) = \frac{P(A)}{P(K)}.$$

Undeniably, any such “subspace” satisfies our Darboux Property.

By $\mathfrak{A}|_K$ we shall understand $\mathfrak{A}(\mathcal{F}_K)$, i.e. the family of all finite subalgebras of \mathcal{F}_K . For any $\mathcal{A} \in \mathfrak{A}$, by $\mathcal{A}|_K \in \mathfrak{A}|_K$ we shall denote $\{A \cap K : A \in \mathcal{A}\}$. For every finite family $\mathcal{G} \subset \mathcal{F}$, by $\sigma(\mathcal{G}) \in \mathfrak{A}$ we shall understand the algebra generated by \mathcal{G} , that is the smallest algebra containing \mathcal{G} .

In the sequel we shall often refer to algebras generated by partitions. This is why the following special notation might prove convenient. We shall write

$$\mathcal{A} = \langle A_1, \dots, A_n \rangle,$$

if $\mathcal{A} = \sigma(\{A_1, \dots, A_n\})$ and $A_1 + \dots + A_n = \Omega$, $A_i \neq \emptyset$, $A_i \in \mathcal{F}$.

Clearly, every algebra $\mathcal{A} \in \mathfrak{A}$ is of shape $\langle A_1, \dots, A_n \rangle$. In fact, we have

$$\{A_i : 1 \leq i \leq n\} = \min_{\subset}(\mathcal{A})$$

where $\min_{\subset}(\mathcal{A})$ denotes the set of minimal elements of $\mathcal{A} \setminus \{\emptyset\}$ partially ordered by the relation \subset (the set of **atoms** of algebra \mathcal{A}).

Notice that, for $\mathcal{A} = \langle A_i \rangle$ and $\mathcal{B} = \langle B_j \rangle$ we have $\sigma(\mathcal{A} \cup \mathcal{B}) = \langle C_1, \dots, C_s \rangle$, where $\{C_1, \dots, C_s\} := \{A_i \cap B_j : A_i \cap B_j \neq \emptyset, 1 \leq i \leq n, 1 \leq j \leq m\}$.

For any $\mathcal{A}, \mathcal{B} \in \mathfrak{A}$ we shall write $\mathcal{A} \perp \mathcal{B}$ whenever these algebras are independent, that is when $P(A \cap B) = P(A)P(B)$ for all $A \in \mathcal{A}$, $B \in \mathcal{B}$. One reason for using this symbol is given by the following observation: “Algebras \mathcal{A} and \mathcal{B} are independent if and only if $\int fg \, dP = 0$ for any simple functions f and g , which are measurable with respect to the corresponding algebras \mathcal{A} and \mathcal{B} , and which have their expected values equal to 0, $\int f \, dP = \int g \, dP = 0$ ”.

Under the representations $\mathcal{A} = \langle A_1, \dots, A_n \rangle$ and $\mathcal{B} = \langle B_1, \dots, B_m \rangle$, the independence $\mathcal{A} \perp \mathcal{B}$ is equivalent to having $P(A_i \cap B_j) = P(A_i)P(B_j)$ for all i, j . We shall also write

$$\mathcal{C} = \mathcal{A} \cdot \mathcal{B},$$

if $\mathcal{A} \perp \mathcal{B}$ and $\mathcal{C} = \sigma(\mathcal{A} \cup \mathcal{B})$.

This paper is devoted to exploring functions of the following kind:

Definition 1. *The mapping $I: \mathfrak{A} \rightarrow \mathbb{R}$ is called an **additive partition entropy** if it satisfies*

$$I(\mathcal{A} \cdot \mathcal{B}) = I(\mathcal{A}) + I(\mathcal{B}),$$

that is if for any $\mathcal{A}, \mathcal{B} \in \mathfrak{A}$ such that $\mathcal{A} \perp \mathcal{B}$ we get $I(\sigma(\mathcal{A} \cup \mathcal{B})) = I(\mathcal{A}) + I(\mathcal{B})$.

Let us state several examples that will play a role in the general description of additive partition entropies. Before we do that, however, consider the following crucial function $L: \mathfrak{A} \rightarrow \mathbb{R}^\Omega$, which to a given algebra $\mathcal{A} = \langle A_1, \dots, A_n \rangle \in \mathfrak{A}$ assigns a simple function

$$L(\mathcal{A}) := \sum_{1 \leq i \leq n} \left(\log \frac{1}{P(A_i)} \right) 1_{A_i}.$$

Evidently, the operator L solves our equation $L(\mathcal{A} \cdot \mathcal{B}) = L(\mathcal{A}) + L(\mathcal{B})$. This leads to the following

Example 1. *Take an arbitrary finitely-additive measure $m: \mathcal{F} \rightarrow \mathbb{R}$, vanishing on sets of P -measure zero. We obtain an additive partition entropy*

$$L_m(\mathcal{A}) := \int L(\mathcal{A}) \, dm = \sum_{1 \leq i \leq n} m(A_i) \log \frac{1}{P(A_i)}.$$

A special case $m = P$, gives the Shannon entropy of \mathcal{A}

$$L_P(\mathcal{A}) = EL(\mathcal{A}) = \sum_{1 \leq i \leq n} P(A_i) \log \frac{1}{P(A_i)}.$$

By taking variance instead of expectation, we arrive at

Example 2. *For any algebra $\mathcal{A} \in \mathfrak{A}$ set*

$$V(\mathcal{A}) := D^2[L(\mathcal{A})] = \int [L(\mathcal{A}) - EL(\mathcal{A})]^2 \, dP.$$

Since $V(\mathcal{A})$ is the variance of the random variable $L(\mathcal{A})$ with respect to probability P , and since for any independent $\mathcal{A}, \mathcal{B} \in \mathfrak{A}$ the random variable $L(\mathcal{A} \cdot \mathcal{B}) = L(\mathcal{A}) + L(\mathcal{B})$ is the sum of independent random variables $L(\mathcal{A})$ and $L(\mathcal{B})$, it follows that the mapping V is an additive partition entropy.

Shannon entropy and Example 2 could be generalised further to any cumulant. In fact, consider the cumulant generating function (cgf) of L

$$t \mapsto \log (E \exp(tL)) = \log \left(\sum_{1 \leq i \leq n} P(A_i)^{1-t} \right).$$

Recall that the cgf of a sum of independent random variables is equal to the sum of cgf's of the respective variables. Therefore, any 'linear functional on the cgf' is an example of an additive partition entropy. This can be the cumulants i.e. the coefficients of the power series representation of cgf of L . It is also convenient to consider the values of cgf (say at $t = 1 - \alpha$), up to a constant factor they are the

Example 3. *Rényi entropies of order $\alpha \neq 1$, namely*

$$R_\alpha(A) := \frac{1}{1-\alpha} \log \left(\sum_{1 \leq i \leq n} P(A_i)^\alpha \right).$$

As explained above, or seen directly, these are additive partition entropies. In fact, this is the most widely known class of examples of an additive entropy. (These entropies were introduced in [21].)

The role of the leading coefficient $1/(1 - \alpha)$ is to ensure that the Rényi entropy tend to Shannon entropy as α tends to 1, (this follows, for instance, from a simple use of de l'Hospitals rule, see also [2]). The case $\alpha = 0$ is somewhat special, too. In this case

Example 4. R_0 takes the shape

$$A \mapsto \log \# \{A_i : P(A_i) > 0\},$$

*and is known as **Hartley entropy**. (It was defined in [16].)*

Our final pair of examples have an important use in the research of extreme cases of complexity theory, (see [26]).

Example 5. *They are the minimum, and the maximum of the simple function $L(A)$:*

$$\begin{aligned} L_{\min}(\langle A_1, \dots, A_n \rangle) &:= \min_{\substack{1 \leq i \leq n \\ P(A_i) \neq 0}} \log \frac{1}{P(A_i)}, \\ L_{\max}(\langle A_1, \dots, A_n \rangle) &:= \max_{\substack{1 \leq i \leq n \\ P(A_i) \neq 0}} \log \frac{1}{P(A_i)}. \end{aligned}$$

Definition 2. *We say that algebras \mathcal{A} and \mathcal{B} have the same measures of atoms, if we have*

$$\mathcal{A} = \langle A_1, \dots, A_n \rangle, \quad \mathcal{B} = \langle B_1, \dots, B_n \rangle$$

*with $P(A_i) = P(B_i)$, for $1 \leq i \leq n$. Also, an additive partition entropy I will be said to **depend solely on the measures of atoms** if I is constant on each family of algebras that have the same measures of atoms.*

This is the same as saying that there is a 'classical' additive entropy H , such that $I = H_P$, i.e. $I(\mathcal{A}) = H(P(A_1), \dots, P(A_n))$.

Remark 1. *The additive partition entropies of Examples 2–5 depend solely on the measures of atoms. In cases when \mathfrak{m} is absolutely continuous with respect to P , (see Definition 4 on page 16) the additive partition entropy $L_{\mathfrak{m}}$*

from Example 1 depends solely on the measures of atoms exactly when $\mathbf{m} = \alpha P$, for some $\alpha \in \mathbb{R}$, *i.e. when it is a multiple of Shannon's entropy.*

Indeed, if $L_{\mathbf{m}}$ depends solely on the measures of atoms then there is a function $f: \mathbb{R} \rightarrow \mathbb{R}$ such that for any $A \in \mathcal{F}$ we have $\mathbf{m}(A) = f(P(A))$. (In fact, if $\mathcal{A} = \langle A, \Omega \setminus A \rangle$, $\mathcal{B} = \langle B, \Omega \setminus B \rangle$, $P(A) = P(B) =: \alpha \neq 1/2$ then we put $x := \mathbf{m}(A)$, $y := \mathbf{m}(B)$ and get

$$(3) \quad \begin{aligned} L_{\mathbf{m}}(\mathcal{A}) &= \left(\log \frac{1}{\alpha} \right) x + \left(\log \frac{1}{1-\alpha} \right) (1-x) = \\ L_{\mathbf{m}}(\mathcal{B}) &= \left(\log \frac{1}{\alpha} \right) y + \left(\log \frac{1}{1-\alpha} \right) (1-y); \end{aligned}$$

from which $x = y$. If $\alpha = 1/2$ break A and B into two pieces of equal measures.) The function f must be additive and continuous at zero, thus linear.

Incidentally, observe that the mapping $\mathbf{m} \mapsto L_{\mathbf{m}}$ is injective. This follows in nearly the same way as equations (3) above.

3. SOME INTRODUCTORY STATEMENTS

From now on, when speaking of an **algebra** or writing \mathcal{A} , \mathcal{B} , etc. we shall mean an algebra in \mathfrak{A} , or in other words an algebra generated by a partition of Ω into a finite number of measurable sets.

For any algebras \mathcal{A} , \mathcal{B} , and $\mathcal{K} = \langle K_1, \dots, K_n \rangle$ we shall write $\mathcal{A} \perp_{\mathcal{K}} \mathcal{B}$ if for every K_i such that $P(K_i) > 0$ the algebra $\mathcal{A}|_{K_i}$ is independent with $\mathcal{B}|_{K_i}$.

Lemma 1. *Let $\mathcal{K} = \langle K_1, \dots, K_n \rangle \subset \mathcal{A}$ and $P(K_i) > 0$, $1 \leq i \leq n$. Then the following conditions are equivalent*

- (1) $\mathcal{A} \perp \mathcal{B}$,
- (2) $\mathcal{B} \perp \mathcal{K}, \mathcal{A} \perp_{\mathcal{K}} \mathcal{B}$.

If these conditions are satisfied then

$$\mathcal{C} = \mathcal{A} \cdot \mathcal{B} \quad \text{is equivalent to} \quad K_i \in \mathcal{C}, \mathcal{C}|_{K_i} = \mathcal{A}|_{K_i} \cdot \mathcal{B}|_{K_i}, \quad 1 \leq i \leq n.$$

Proof. We leave to the reader the proof that 1. and 2. are equivalent.

To show the second equivalence we need the following for $\mathcal{K} \subset \mathcal{C}$

$$(4) \quad \mathcal{C} = \sigma(\mathcal{A} \cup \mathcal{B}) \iff \mathcal{C}|_{K_i} = \sigma(\mathcal{A}|_{K_i} \cup \mathcal{B}|_{K_i}) \quad (1 \leq i \leq n).$$

Since $\mathcal{K} \subset \mathcal{C}$ we have

$$(5) \quad \min_{\mathcal{C}}(\mathcal{C}) = \sum_i \min_{\mathcal{C}}(\mathcal{C}|_{K_i}).$$

Similarly from $\mathcal{K} \subset \sigma(\mathcal{A} \cup \mathcal{B})$ we obtain

$$(6) \quad \min_{\mathcal{C}}(\sigma(\mathcal{A} \cup \mathcal{B})) = \sum_i \min_{\mathcal{C}}(\sigma(\mathcal{A} \cup \mathcal{B})|_{K_i}) = \sum_i \min_{\mathcal{C}}(\sigma(\mathcal{A}|_{K_i} \cup \mathcal{B}|_{K_i})).$$

From (5) and (6) we obtain (4). □

Lemma 2. *For any algebra \mathcal{A} and any numbers $c_1, \dots, c_k \geq 0$ such that $\sum_{1 \leq j \leq k} c_j = 1$ there is an algebra $\mathcal{C} = \langle C_1, \dots, C_k \rangle$ satisfying $\mathcal{C} \perp \mathcal{A}$, $P(C_i) = c_i$.*

Proof. This follows easily from the Darboux property. \square

Corollary 1. *For any algebras $\mathcal{A}_1, \dots, \mathcal{A}_n$ and any nonnegative numbers c_1, \dots, c_k such that $\sum_{1 \leq j \leq k} c_j = 1$ there is an algebra $\mathcal{C} = \langle C_1, \dots, C_k \rangle$ satisfying $\mathcal{C} \perp \mathcal{A}_i$, $1 \leq i \leq n$, $P(C_i) = c_i$.*

Proof. Consider $\mathcal{A} = \sigma(\mathcal{A}_1, \dots, \mathcal{A}_n)$. \square

Remark 2. *We will often use this Corollary in the following way. For any algebras \mathcal{A} and \mathcal{B} having the same measures of atoms there is an algebra \mathcal{C} with the same measures of atoms such that*

$$\mathcal{C} \perp \mathcal{A}, \quad \mathcal{C} \perp \mathcal{B}.$$

Fix any additive partition entropy I .

Proposition 1. *If for some set Z , $P(Z) = 0$ we have*

$$\mathcal{A}|_{\Omega \setminus Z} = \mathcal{B}|_{\Omega \setminus Z},$$

then

$$I(\mathcal{A}) = I(\mathcal{B}).$$

Proof. We can easily find representations $\mathcal{A} = \langle \mathcal{A}_1, \dots, \mathcal{A}_n, \dots, \mathcal{A}_s \rangle$ and $\mathcal{B} = \langle \mathcal{B}_1, \dots, \mathcal{B}_n, \dots, \mathcal{B}_t \rangle$ for which the following equalities hold

$$P(\mathcal{A}_i \triangle \mathcal{B}_i) = 0 \quad (1 \leq i \leq n), \quad P(\mathcal{A}_{n+k}) = 0, \quad P(\mathcal{B}_{n+l}) = 0 \quad (k, l \geq 1).$$

Observe that any algebra \mathcal{C} generated by sets of measure 0 is independent with any other algebra from \mathfrak{A} . The algebra $\mathcal{C} := \sigma(\{\mathcal{A}_i \triangle \mathcal{B}_i, \mathcal{A}_{n+k}, \mathcal{B}_{n+l}\})$ is of such a shape and $\mathcal{A} \cdot \mathcal{C} = \mathcal{B} \cdot \mathcal{C}$. \square

Proposition 2. *Consider algebras $\mathcal{A} = \langle \mathcal{A}_1, \dots, \mathcal{A}_k, \mathcal{A}_{k+1}, \dots, \mathcal{A}_n \rangle$ and $\mathcal{A}' = \langle \mathcal{A}'_1, \dots, \mathcal{A}'_k, \mathcal{A}_{k+1}, \dots, \mathcal{A}_n \rangle$ with*

$$P(\mathcal{A}_1) = \dots = P(\mathcal{A}_k) = P(\mathcal{A}'_1) = \dots = P(\mathcal{A}'_k).$$

Then

$$I(\mathcal{A}) = I(\mathcal{A}').$$

Proof. Put $K := \mathcal{A}_1 + \dots + \mathcal{A}_k$. Since the case $P(K) = 0$ is handled by the previous Proposition we assume $P(K) > 0$. By the same Proposition we shall assume $K = \Omega$ whenever $P(K) = 1$. Corollary 1 (see Remark 2) when applied to $\mathfrak{A}|_K$ lets us suppose that $\mathcal{A}|_K \perp \mathcal{A}'|_K$.

We shall find \mathcal{B} such that $\mathcal{A} \cdot \mathcal{B} = \mathcal{A}' \cdot \mathcal{B}$. Firstly, define an algebra $\mathcal{B}_K = \langle \mathcal{B}_K^{(1)}, \dots, \mathcal{B}_K^{(k)} \rangle \in \mathfrak{A}|_K$ by

$$\mathcal{B}_K^{(i)} := \sum_{\substack{q-p \equiv i \pmod{k} \\ 1 \leq p, q \leq k}} \mathcal{A}_p \cap \mathcal{A}'_q \quad (1 \leq i \leq k).$$

Then $\mathcal{A}|_K \cdot \mathcal{B}_K = \mathcal{A}|_K \cdot \mathcal{A}'|_K = \mathcal{A}'|_K \cdot \mathcal{B}_K$.

If $K \neq \Omega$ then Lemma 2 allows us to find an algebra

$$\mathcal{B}_{\Omega \setminus K} = \langle \mathcal{B}_{\Omega \setminus K}^{(1)}, \dots, \mathcal{B}_{\Omega \setminus K}^{(k)} \rangle \in \mathfrak{A}|_{\Omega \setminus K}$$

that has the same measures of atoms as \mathcal{B}_K and satisfies $\mathcal{B}_{\Omega \setminus K} \perp \mathcal{A}|_{\Omega \setminus K}$. Consider

$$\mathcal{B} := \langle \mathcal{B}_K^{(1)} \cup \mathcal{B}_{\Omega \setminus K}^{(1)}, \dots, \mathcal{B}_K^{(k)} \cup \mathcal{B}_{\Omega \setminus K}^{(k)} \rangle \in \mathfrak{A}.$$

By Lemma 1 we obtain the equality $\mathcal{A} \cdot \mathcal{B} = \mathcal{A}' \cdot \mathcal{B}$. \square

Remark 3. *As we have seen, all our examples can be made to rely on the function L defined just before Example 1. This is by no coincidence. Propositions 1 and 2 mean that if $L(\mathcal{A})$ and $L(\mathcal{B})$ differ on a set of measure 0 then $I(\mathcal{A}) = I(\mathcal{B})$; in particular every additive partition entropy I factors through L , i.e. $I = J \circ L$, where*

$$J(\phi + \psi) = J(\phi) + J(\psi) \quad \text{with } \phi, \psi \in \text{Im}(L) \text{ and } \phi, \psi \text{ independent.}$$

Moreover it seems quite plausible that this condition is satisfied for $\phi, \psi, \phi + \psi \in \text{Im}(L)$. In this paper, we don't pursue this approach any further.

For any disjoint sets $V, W \in \mathcal{F}$ define a nonempty family \mathfrak{F}_{VW} of algebras \mathcal{A} which satisfy $V \subset A_1$, $W \subset A_2$ for some representation $\langle A_1, \dots, A_n \rangle$ of algebra \mathcal{A} . Also define an operation $\mathcal{T}_{VW}: \mathfrak{F}_{VW} \rightarrow \mathfrak{F}_{VW}$ by

$$\mathcal{T}_{VW} \langle A_1, \dots, A_n \rangle = \langle (A_1 \setminus V) \cup W, (A_2 \setminus W) \cup V, A_3, \dots, A_n \rangle,$$

when $V \subset A_1$, $W \subset A_2$.

Lemma 3. *Consider a pair of disjoint sets V, W with $P(V) = P(W)$, and algebras $\mathcal{A}, \mathcal{B} \in \mathfrak{F}_{VW}$. Whenever*

$$\mathcal{C} = \mathcal{A} \cdot \mathcal{B},$$

we also have

$$\mathcal{T}_{VW} \mathcal{C} = \mathcal{T}_{VW} \mathcal{A} \cdot \mathcal{T}_{VW} \mathcal{B}.$$

Proof. We can set

$$\mathcal{A} = \langle A_1, \dots, A_n \rangle, \quad \mathcal{B} = \langle B_1, \dots, B_m \rangle, \quad \mathcal{C} = \langle C_1, \dots, C_p \rangle,$$

with $V \subset A_1 \cap B_1 = C_1$, $W \subset A_2 \cap B_2 = C_2$. Put $A'_i := A_i \Delta V \Delta W$, $B'_j := B_j \Delta V \Delta W$ and $C'_k := C_k \Delta V \Delta W$ for $i, j, k = 1, 2$, and also $A'_i = A_i$, $B'_j = B_j$, $C'_k = C_k$ for $3 \leq i \leq n$, $3 \leq j \leq m$, $3 \leq k \leq p$. Then $\langle A'_1, \dots, A'_n \rangle = \mathcal{T}_{VW} \mathcal{A}$, $\langle B'_1, \dots, B'_m \rangle = \mathcal{T}_{VW} \mathcal{B}$, $\langle C'_1, \dots, C'_p \rangle = \mathcal{T}_{VW} \mathcal{C}$ and $\langle C'_1, \dots, C'_p \rangle = \sigma(A'_1, \dots, A'_n, B'_1, \dots, B'_m)$.

It suffices to show that

$$(7) \quad \begin{aligned} P(A'_i) &= P(A_i), & P(B'_j) &= P(B_j), \\ P(A'_i \cap B'_j) &= P(A_i \cap B_j). \end{aligned}$$

When $i = j = 1, 2$ this follows from equalities $P(V) = P(W)$ and inclusions $V \subset A_1 \cap B_1$, $W \subset A_2 \cap B_2$. For all remaining pairs (i, j) , $i = 1, \dots, n$,

$j = 1, \dots, m$ the last of equalities in (7) is also clear, and we even have $A'_i \cap B'_j = A_i \cap B_j$. \square

For any $\lambda > 0$ put

$$\varepsilon(\lambda) := \min \left(1/(1 + \lambda), 1/(1 + \lambda^{-1}) \right).$$

This notation has the following sense — if we divide Ω into two sets having their quotient of measures equal to λ then $\varepsilon(\lambda)$ will be the measure of the smaller of them.

Whenever $V \cap W = \emptyset$ and $\lambda > 0$ write

$$\mathfrak{F}_{VW}^\lambda := \left\{ \mathcal{A} \in \mathfrak{A}: \frac{P(A_2)}{P(A_1)} = \lambda, V \subset A_1, W \subset A_2 \right. \\ \left. \text{for some representation } \mathcal{A} = \langle A_1, \dots, A_n \rangle \right\}.$$

Lemma 4. *For any $\lambda > 0$ we have what follows:*

A. *For any pair of disjoint sets V, W such that $P(V), P(W) \leq \varepsilon(\lambda)$ there is an algebra $\mathcal{A} = \langle A_1, A_2 \rangle \in \mathfrak{F}_{VW}^\lambda$.*

B. *For any algebra $\mathcal{A} = \langle A_1, \dots, A_n \rangle$, $n \geq 2$ with $P(A_2)/P(A_1) = \lambda$, a number $\kappa > 0$ and sets $V \subset A_1$, $W \subset A_2$ that satisfy*

$$P(V), P(W) \leq \varepsilon(\kappa)\varepsilon(\lambda)P(A_1 + A_2)$$

there is an algebra $\mathcal{B} = \langle B_1, B_2 \rangle$ with the property that $\mathcal{A} \perp \mathcal{B}$, $\mathcal{B} \in \mathfrak{F}_{VW}^\kappa$.

C. *For any disjoint V, W and $\kappa > 0$ such that $P(V), P(W) \leq \varepsilon(\kappa)\varepsilon(\lambda)$, there exist algebras $\mathcal{A} = \langle A_1, A_2 \rangle$, $\mathcal{B} = \langle B_1, B_2 \rangle$ such that*

$$\mathcal{A} \in \mathfrak{F}_{VW}^\lambda, \quad \mathcal{B} \in \mathfrak{F}_{VW}^\kappa, \quad \text{and} \quad \mathcal{A} \perp \mathcal{B}.$$

D. *If $\mathcal{A} \in \mathfrak{F}_{VW}^\lambda$, $\mathcal{B} \in \mathfrak{F}_{VW}^\kappa$ and $\mathcal{C} = \mathcal{A} \cdot \mathcal{B}$ then $\mathcal{C} \in \mathfrak{F}_{VW}^{\lambda\kappa}$.*

Proof. **A.** If $P(V), P(W) \leq \varepsilon(\lambda)$ then by the Darboux property there is a partition $\Omega = A_1 + A_2$ such that $P(A_1) = 1/(1 + \lambda)$ and $P(A_2) = 1/(1 + \lambda^{-1})$ with $V \subset A_1$, $W \subset A_2$. Then $P(A_2)/P(A_1) = \lambda$.

B. By the assumptions $P|_{A_1}(V), P|_{A_2}(W) \leq \varepsilon(\kappa)$. According to **A.** there exist sets $C_i^{(j)} \in \mathfrak{A}|_{A_i}$, $1 \leq i \leq n$, $j = 1, 2$ such that $A_i = C_i^{(1)} + C_i^{(2)}$, $P(C_i^{(2)})/P(C_i^{(1)}) = \kappa$, and $V \subset C_1^{(1)}$, $W \subset C_2^{(2)}$. We finish by writing

$$B_j := C_1^{(j)} + \dots + C_n^{(j)}.$$

C. follows from **A.** and **B.**, whereas **D.** is obvious. \square

The following lemma is utterly straightforward.

Lemma 5. *Whenever $\mathcal{A} \in \mathfrak{F}_{VW}^\lambda$, we have $\mathcal{A} \in \mathfrak{F}_{V_i W_i}^\lambda$, $1 \leq i \leq n$ and*

$$\mathcal{T}_{VW}\mathcal{A} = \mathcal{T}_{V_1 W_1} \cdots \mathcal{T}_{V_n W_n} \mathcal{A}$$

for any $V = V_1 + \dots + V_n$, $W = W_1 + \dots + W_n$.

Remark 4. *It follows from Proposition 2 that if $\mathcal{A} \in \mathfrak{F}_{VW}^1$ then*

$$I(\mathcal{T}_{VW}\mathcal{A}) = I(\mathcal{A}).$$

Proposition 3. *Consider $\lambda > 0$ and a pair of disjoint sets V, W with $P(V) = P(W)$.*

If $\mathcal{A}, \mathcal{B} \in \mathfrak{F}_{VW}^\lambda$ then

$$I(\mathcal{T}_{VW}\mathcal{A}) - I(\mathcal{A}) = I(\mathcal{T}_{VW}\mathcal{B}) - I(\mathcal{B}).$$

Proof. Fix algebras \mathcal{A} and \mathcal{B} .

By Proposition 1 we can assume that $P(V) > 0$; then for some $\kappa > 0$ we get $\sigma(\mathcal{A} \cup \mathcal{B}) \in \mathfrak{F}_{VW}^\kappa$.

Suppose for the moment that we also have

$$P(V) = P(W) \leq \varepsilon(1/\lambda) \cdot \varepsilon(\kappa) \cdot P(A_1 \cap B_1 + A_2 \cap B_2),$$

where A_i, B_i are such atoms of algebras \mathcal{A} and \mathcal{B} that $V \subset A_1 \cap B_1$ and $W \subset A_2 \cap B_2$.

With the use of Lemma 4B we find an algebra $\mathcal{C} = \langle C_1, C_2 \rangle \in \mathfrak{F}_{VW}^{1/\lambda}$ such that $\sigma(\mathcal{A} \cup \mathcal{B}) \perp \mathcal{C}$. Then, using Lemma 3, Lemma 4D and the Remark above we get the equalities $I(\mathcal{A} \cdot \mathcal{C}) = I(\mathcal{T}_{VW}\mathcal{A} \cdot \mathcal{T}_{VW}\mathcal{C})$ and $I(\mathcal{B} \cdot \mathcal{C}) = I(\mathcal{T}_{VW}\mathcal{B} \cdot \mathcal{T}_{VW}\mathcal{C})$. We are done.

In the general case divide the sets V and W into the same number of pieces of equal measure, the measure being bound by

$$\varepsilon(1/\lambda) \cdot \varepsilon(\kappa) \cdot P(A_1 \cap B_1 + A_2 \cap B_2).$$

Subsequently apply Lemma 5 for $\mathcal{A}, \mathcal{B} \in \mathfrak{F}_{VW}^\lambda$ and $\sigma(\mathcal{A} \cup \mathcal{B}) \in \mathfrak{F}_{VW}^\kappa$ to the already derived instance of this Proposition where we have constraints on the size of V and W . \square

Proposition 4. *For any sets V, W with $P(V) = P(W)$ and any $\lambda > 0$ there is a unique $\Delta(V, W, \lambda)$ such that the following conditions are satisfied:*

(1) *Whenever $V \cap W = \emptyset$ and there exists $\mathcal{A} \in \mathfrak{F}_{VW}^\lambda$ we have*

$$(8) \quad \Delta(V, W, \lambda) = I(\mathcal{T}_{VW}\mathcal{A}) - I(\mathcal{A}),$$

(2) *For any sets U, V and W of the same measure P we have*

$$\Delta(U, W, \lambda) = \Delta(U, V, \lambda) + \Delta(V, W, \lambda).$$

(3) *For $V = V_1 + V_2$, $W = W_1 + W_2$, $P(V_i) = P(W_i)$, $i = 1, 2$ we have*

$$\Delta(V, W, \lambda) = \Delta(V_1, W_1, \lambda) + \Delta(V_2, W_2, \lambda).$$

(4) *For $\kappa > 0$ we have*

$$\Delta(V, W, \kappa\lambda) = \Delta(V, W, \kappa) + \Delta(V, W, \lambda).$$

Proof. Notice at first that in the case when the family $\mathfrak{F}_{VW}^\lambda$ is nonempty, Proposition 3 proves that the right hand side of equality (8) does not depend on the algebra $\mathcal{A} \in \mathfrak{F}_{VW}^\lambda$, i.e. that formula (8) defines the quantity $\Delta(V, W, \lambda)$ well. Observe also that by Lemma 4A the family $\mathfrak{F}_{VW}^\lambda$ is nonempty when the sets V, W are disjoint and satisfy $P(V) = P(W) < \varepsilon(\lambda)$.

The proof is made of three parts.

I. Assume that $\mathfrak{F}_{VW}^\lambda \neq \emptyset$ and define for now Δ simply by (8). In particular we assume that V, W are disjoint. Using just defined Δ 's we shall prove

property 2 when $P(V) = P(W) < (1/2)\varepsilon(\lambda)$, property 3 assuming that $P(V) = P(W) < \varepsilon(\lambda)$ and property 4 if $P(V) = P(W) < \varepsilon(\kappa)\varepsilon(\lambda)$.

Indeed, under these assumptions we shall show property 2 by choosing an algebra \mathcal{A} in such a way that $U \in A_1$, $V \cup W \in A_2$, where A_2 is the atom of \mathcal{A} with λ times bigger P -measure than A_1 , and noting that

$$\mathcal{A} \in \mathfrak{F}_{UW}^\lambda, \quad \mathcal{A} \in \mathfrak{F}_{UV}^\lambda, \quad \mathcal{T}_{UV}\mathcal{A} \in \mathfrak{F}_{VW}^\lambda, \quad (\mathcal{T}_{VW} \circ \mathcal{T}_{UV})\mathcal{A} = \mathcal{T}_{UW}\mathcal{A}.$$

Then

$$\begin{aligned} \Delta(U, W, \lambda) &= I(\mathcal{T}_{UW}\mathcal{A}) - I(\mathcal{A}) \\ &= I((\mathcal{T}_{VW} \circ \mathcal{T}_{UV})\mathcal{A}) - I(\mathcal{T}_{UV}\mathcal{A}) + I(\mathcal{T}_{UV}\mathcal{A}) - I(\mathcal{A}) \\ &= \Delta(V, W, \lambda) + \Delta(U, V, \lambda). \end{aligned}$$

In order to show property 3 we select $\mathcal{A} \in \mathfrak{F}_{VW}^\lambda$ and obtain

$$\begin{aligned} \Delta(V, W, \lambda) &= I(\mathcal{T}_{V_2W_2} \circ \mathcal{T}_{V_1W_1}\mathcal{A}) - I(\mathcal{A}) \\ &= I(\mathcal{T}_{V_2W_2} \circ \mathcal{T}_{V_1W_1}\mathcal{A}) - I(\mathcal{T}_{V_1W_1}\mathcal{A}) + I(\mathcal{T}_{V_1W_1}\mathcal{A}) - I(\mathcal{A}) \\ &= \Delta(V_2, W_2, \lambda) + \Delta(V_1, W_1, \lambda). \end{aligned}$$

It remains to prove property 4. Observe that by Lemma 4C there exist $\mathcal{A} \in \mathfrak{F}_{VW}^\lambda$, $\mathcal{B} \in \mathfrak{F}_{VW}^\kappa$ such that $\mathcal{A} \perp \mathcal{B}$. What is more, by Lemma 4D we have $\mathcal{A} \cdot \mathcal{B} \in \mathfrak{F}_{VW}^{\lambda \cdot \kappa}$; now using Lemma 3 we get

$$\begin{aligned} \Delta(V, W, \lambda \kappa) &= I(\mathcal{T}_{VW}(\mathcal{A} \cdot \mathcal{B})) - I(\mathcal{A} \cdot \mathcal{B}) \\ &= I(\mathcal{T}_{VW}\mathcal{A}) - I(\mathcal{A}) + I(\mathcal{T}_{VW}\mathcal{B}) - I(\mathcal{B}) \\ &= \Delta(V, W, \lambda) + \Delta(V, W, \kappa). \end{aligned}$$

Notice also that if $P(V) < \varepsilon(\lambda)$ then

$$\Delta(V, W, \lambda) = -\Delta(W, V, \lambda),$$

which we shall use in the next step of the proof.

II. We drop now the assumption that the sets V and W be disjoint, that is we consider the case when $\mathfrak{F}_{V \setminus W, W \setminus V}^\lambda \neq \emptyset$ (e.g. when $P(V) = P(W) < \varepsilon(\lambda)$), and define Δ by

$$\Delta(V, W, \lambda) := \Delta(V \setminus W, W \setminus V, \lambda)$$

By supposing, if neccessary, that we have $P(V) < 1/4$ in addition to the assumptions of part **I**, we shall show the required properties 2 and 3 (property 4 is obvious), without assuming that V and W are disjoint.

In order to show these properties we note at first that for any U , V and W having the required properties and for any X such that $P(X) = P(V)$ and $X \cap (U \cup V \cup W) = \emptyset$ we have

$$\Delta(V, W, \lambda) = \Delta(V, X, \lambda) + \Delta(X, W, \lambda)$$

Indeed set $V' := V \setminus W$, $W' := W \setminus V$, $A = V \cap W$ then divide the set X into two parts — X' of measure $P(X') = P(V')$ and B of measure $P(B) = P(A)$.

Using step **I.**, we obtain:

$$\begin{aligned}
\Delta(V, W, \lambda) &= \Delta(V', W', \lambda) \\
&= \Delta(V', X', \lambda) + \Delta(X', W', \lambda) \\
&= \Delta(V', X', \lambda) + \Delta(A, B, \lambda) + \Delta(X', W', \lambda) + \Delta(B, A, \lambda) \\
&= \Delta(V, X, \lambda) + \Delta(X, W, \lambda).
\end{aligned}$$

From the obtained equality we arrive at property 2:

$$\begin{aligned}
\Delta(U, W, \lambda) &= \Delta(U, X, \lambda) + \Delta(X, W, \lambda) \\
&= \Delta(U, X, \lambda) + \Delta(X, V, \lambda) + \Delta(V, X, \lambda) + \Delta(X, W, \lambda) \\
&= \Delta(U, V, \lambda) + \Delta(V, W, \lambda),
\end{aligned}$$

and also, by a division of the set X into the parts X_1, X_2 , with their measures equal to $P(V_1)$ and $P(V_2)$, respectively we arrive at property 3:

$$\begin{aligned}
\Delta(V, W, \lambda) &= \Delta(V, X, \lambda) + \Delta(X, W, \lambda) \\
&= \Delta(V_1, X_1, \lambda) + \Delta(V_2, X_2, \lambda) + \Delta(X_1, W_1, \lambda) + \Delta(X_2, W_2, \lambda) \\
&= \Delta(V_1, W_1, \lambda) + \Delta(V_2, W_2, \lambda).
\end{aligned}$$

III. In the general case ($P(V) = P(W)$, $\lambda > 0$) the number $\Delta(V, W, \lambda)$ can be uniquely defined by

$$\Delta(V, W, \lambda) := \sum_{1 \leq i \leq k} \Delta(V_i, W_i, \lambda),$$

where $V = V_1 + \dots + V_k$, $W = W_1 + \dots + W_k$, $P(V_i) = P(W_i) < \varepsilon(\lambda)$; and where $\Delta(V_i, W_i, \lambda)$ is defined as in part **II.**

We shall show now that the number $\Delta(V, W, \lambda)$ is well-defined in this way. For any partitions $V = V_1 + \dots + V_k$, $W = W_1 + \dots + W_k$ and $V = V'_1 + \dots + V'_l$, $W = W'_1 + \dots + W'_l$ satisfying respectively $P(V_i) = P(W_i) < \varepsilon(\lambda)$, $1 \leq i \leq k$ and $P(V'_j) = P(W'_j) < \varepsilon(\lambda)$, $1 \leq j \leq l$, we choose another pair of partitions $V = V''_1 + \dots + V''_m$, $W = W''_1 + \dots + W''_m$ in such a way that $P(V''_i) = P(W''_i) < \varepsilon(\lambda)$, and $\langle V''_1, \dots, V''_m \rangle$ is independent with $\langle V_1, \dots, V_k \rangle$ and $\langle V'_1, \dots, V'_l \rangle$ in the space $(V, \mathcal{F}_V, P|_V)$, and also that $\langle W''_1, \dots, W''_m \rangle$ is independent with $\langle W_1, \dots, W_k \rangle$ and $\langle W'_1, \dots, W'_l \rangle$ in the space $(W, \mathcal{F}_W, P|_W)$. Then

$$\begin{aligned}
P(V_i \cap V''_j) &= P(W_i \cap W''_j) \\
P(V'_i \cap V''_j) &= P(W'_i \cap W''_j).
\end{aligned}$$

Hence

$$\sum_{1 \leq i \leq k} \Delta(V_i, W_i, \lambda) = \sum_{\substack{1 \leq i \leq k \\ 1 \leq j \leq m}} \Delta(V_i \cap V''_j, W_i \cap W''_j, \lambda) = \sum_{1 \leq j \leq m} \Delta(V''_j, W''_j, \lambda).$$

In the same way

$$\sum_{1 \leq i \leq l} \Delta(V'_i, W'_i, \lambda) = \sum_{1 \leq j \leq m} \Delta(V''_j, W''_j, \lambda).$$

It is easy to see that Properties 2–4 are satisfied in all their generality. \square

Remark 5. *It is worthy to note that Properties 2 and 3 entail formulas*

$$\begin{aligned}\Delta(V, W, \lambda) &= -\Delta(W, V, \lambda), \\ \Delta(V, W, \lambda) &= \Delta(V \setminus W, W \setminus V, \lambda);\end{aligned}$$

moreover, according to Proposition 1, if $P(V \triangle V') = P(W \triangle W') = 0$, then

$$\Delta(V, W, \lambda) = \Delta(V', W', \lambda).$$

4. THE MAIN RESULT

In this section we provide a description of continuous additive partition entropies. However, no notion of continuity has been developed as yet. We would like our notion to be as weak as possible, with many continuous partition entropies. Yet, at the same time we want partition entropies H_p to be continuous exactly when the corresponding classical entropies H are continuous. Here, a subtle distinction should be made. There are two natural definitions of continuity for a classical entropy, depending on whether we consider nonnegative or just positive probabilities. The same phenomenon arises in the context of partition entropies.

We shall consider a pseudometric $d_{\mathcal{F}}$ in algebra \mathcal{F} defined by

$$d_{\mathcal{F}}(A, B) = P(A \triangle B).$$

By analogy with the classical case we want to consider two topologies in the family \mathfrak{A} both introduced as the richest topology so that, in the first case, the mappings “similar to the following ones”

$$\{A \in \mathcal{F} : 0 \leq P(A) \leq 1\} \ni A \mapsto \langle A, \Omega \setminus A \rangle \in \mathfrak{A}$$

and, in the second case, the following ones

$$\{A \in \mathcal{F} : 0 < P(A) < 1\} \ni A \mapsto \langle A, \Omega \setminus A \rangle \in \mathfrak{A}$$

are continuous. To do so we define the closed-domain topology in \mathfrak{A} by the following pseudometric:

$$d(\mathcal{A}, \mathcal{B}) := \inf \{P(Z) : \mathcal{A}|_{\Omega \setminus Z} = \mathcal{B}|_{\Omega \setminus Z}\},$$

and the stronger open-domain topology by the following one

$$D(\mathcal{A}, \mathcal{B}) := \inf \{P(Z) : \mathcal{A}|_{\Omega \setminus Z} = \mathcal{B}|_{\Omega \setminus Z}\} + |N(\mathcal{A}) - N(\mathcal{B})|,$$

where $N(\mathcal{A})$ denotes the number of atoms of algebra \mathcal{A} with nonvanishing measure.

Definition 3. *A function $I : \mathfrak{A} \rightarrow \mathbb{R}$ is said to be **closed-domain continuous** if it is continuous in metric d and **open-domain continuous** or simply **continuous** if it is continuous in metric D .*

We consider closed-domain continuity to be too restrictive, with many entropies, like some L_m and Hartley entropy not being closed-domain continuous. Open-domain continuity on the other hand, plays nicely with our examples and our theory. To see that let us first make precise the following

Definition 4. A finitely-additive set function $m: \mathcal{F} \rightarrow \mathbb{R}$ is said to be **absolutely continuous** with respect to measure P ($m \ll P$) if it is continuous in pseudometric $d_{\mathcal{F}}$, or equivalently if for any $\varepsilon > 0$ there is a $\delta > 0$ such that if we have $P(A) < \delta$ then we also have $m(A) < \varepsilon$.

Recall that in case of finitely-additive measures the vanishing of m on sets of P -measure 0 does not imply the absolute continuity of m . Now we have

Remark 6. Additive partition entropy L_m from Example 1 is open-domain continuous, when m is absolutely continuous with respect to P . (As it will follow from Theorem 1 and Remark 1 there are no other continuous additive partition entropies of this shape.)

We shall show the continuity of L_m in case when $m \ll P$. Fix algebra $\mathcal{A} = \langle A_1, \dots, A_n \rangle$ and $0 < \varepsilon < 1$. There is a $0 < \delta < 1$ such that if

$$P(A_i \triangle B) < \delta, \quad 1 \leq i \leq n,$$

then

$$|\log P(A_i) - \log P(B)| < \varepsilon \quad \text{and} \quad |m(A_i) - m(B)| < \varepsilon.$$

If now $d(\mathcal{A}, \mathcal{B}) < \delta < 1$ then the algebras \mathcal{A} and \mathcal{B} have the same number of atoms of nonzero measure. We can assume that $\mathcal{B} = \langle B_1, \dots, B_n \rangle$ and $P(A_i), P(B_i) > 0$, and also that $P(A_i \triangle B_i) < \delta$. Then

$$\begin{aligned} |L_m(\mathcal{A}) - L_m(\mathcal{B})| &= \left| \sum m(A_i) \log \frac{1}{P(A_i)} - \sum m(B_i) \log \frac{1}{P(B_i)} \right| \\ &\leq \sum |m(A_i)| \left| \log \frac{1}{P(A_i)} - \log \frac{1}{P(B_i)} \right| + \\ &\quad \sum |(m(A_i) - m(B_i)) \log \frac{1}{P(B_i)}| \\ &< \sum |m(A_i)| \varepsilon + \sum \varepsilon \log \frac{1}{P(B_i)} \\ &< \varepsilon \sum |m(A_i)| + \varepsilon \sum \left(\varepsilon + \log \frac{1}{P(A_i)} \right) \\ &< \varepsilon \cdot \text{const}(\mathcal{A}). \end{aligned}$$

Remark 7. Given a 'classical' additive entropy H , the corresponding partition entropy H_P is continuous if and only if H is continuous on the open domain, that is if each function

$$H|_{\{(p_1, \dots, p_n): p_i > 0, \sum p_i = 1\}}$$

is continuous. H_P is closed-domain continuous iff H is continuous everywhere. It follows that, entropies of Examples 2–5 are continuous.⁵

⁵Please note that Hartley entropy, and L_{\min} are not closed-domain continuous.

It turns out that in the derivation of our main theorem (Theorem 1) we can do with weaker concepts than that of open-domain continuity. In fact, we do not need the continuity on the full family \mathfrak{A} . It will be sufficient to assume that I is continuous on the family of algebras with 2 atoms. Let for that matter \mathfrak{A}_2 denote the family

$$\mathfrak{A}_2 := \{\langle A, B \rangle : 0 < P(A) < 1\}.$$

Definition 5. We say that I is **continuous on \mathfrak{A}_2** if the restriction $I_{\mathfrak{A}_2}$ of I to \mathfrak{A}_2 is continuous, that is if for any sequence of sets $A, A_1, A_2, \dots \in \mathcal{F}$ with $0 < P(A) < 1$ such that $P(A \triangle A_n) \rightarrow 0$ we have

$$I(\langle A_n, \Omega \setminus A_n \rangle) \rightarrow I(\langle A, \Omega \setminus A \rangle).$$

Proposition 5. If the additive partition entropy I is continuous on \mathfrak{A}_2 then for any sets V, W such that $P(V) = P(W)$ there is a $\Delta(V, W)$ satisfying

$$\Delta(V, W, \lambda) = \Delta(V, W) \cdot \log \lambda \quad (\lambda > 0).$$

What is more Δ has the following properties:

(1) For any sets U, V and W of the same measure

$$\Delta(U, W) = \Delta(U, V) + \Delta(V, W).$$

(2) For $V = V_1 + V_2$, $W = W_1 + W_2$, $P(V_i) = P(W_i)$, $i = 1, 2$ we have

$$\Delta(V, W) = \Delta(V_1, W_1) + \Delta(V_2, W_2).$$

(3) $\Delta(\cdot, \cdot)$ is **uniformly** continuous (in the topology induced from the one in \mathcal{F} .)

Proof. By $\Delta(V, W, \lambda) = \Delta(V \setminus W, W \setminus V, \lambda)$ and property 3 of Proposition 4, in order to prove the first part of Proposition, we can assume that the sets V and W are disjoint and that their measures are smaller than $1/2$.

At first, we shall show that the mapping $\lambda \mapsto \Delta(V, W, \lambda)$ is continuous at $\lambda = 1$. Indeed, consider the following open neighbourhood of the point $1 \in \mathbb{R}$: $G := \{\lambda : P(V) < \varepsilon(\lambda)\}$. Consider also any sequence $(\lambda_i)_{i \geq 1}$ such that $\lambda_i \in G$ and $\lambda_i \rightarrow 1$. By the Darboux property (see also Lemma 4A) we shall find algebras $\mathcal{A}_i = \langle A_1^i, A_2^i \rangle \in \mathfrak{F}_{VW}^{\lambda_i}$ and $\mathcal{A} = \langle A_1, A_2 \rangle \in \mathfrak{F}_{VW}^1$ such that $P(A_1^i \triangle A_1) \rightarrow 0$ when $i \rightarrow \infty$. Then, from the continuity of I we get

$$I(\mathcal{A}_i) \rightarrow I(\mathcal{A}),$$

$$I(\mathcal{T}_{VW}\mathcal{A}_i) \rightarrow I(\mathcal{T}_{VW}\mathcal{A}),$$

and consequently $\Delta(V, W, \lambda_i) \rightarrow \Delta(V, W, 1)$.

Next by property 4 of Proposition 4 and the fact that $\Delta(V, W, \cdot)$ is continuous at $\lambda = 1$ we see that there is a constant $\alpha \in \mathbb{R}$ with

$$\Delta(V, W, \lambda) = \alpha \log \lambda \quad (\lambda > 0).$$

Put $\Delta(V, W) := \alpha$.

Property 1 and property 2 follow from the corresponding properties in Proposition 4.

As it follows from property 2, **to get uniform continuity we only need to show that Δ is continuous** at $V = W = \emptyset$; i.e. it suffices to show that for any $\varepsilon > 0$ we can find $\delta > 0$ such that if $P(V) = P(W) < \delta$ then $|\Delta(V, W)| < \varepsilon$.

Fix $\varepsilon > 0$. Using the continuity of the function I at the point $\langle K_1, K_2 \rangle$, for any fixed partition $P(K_1) = 1/3$, $P(K_2) = 2/3$ of the space Ω we shall find $\delta > 0$ such that for any algebras $\langle L_1, L_2 \rangle$ which satisfy $P(K_1 \Delta L_1) < 4\delta$ (and $P(L_1) = 1/3$) we have $|I(\langle K_1, K_2 \rangle) - I(\langle L_1, L_2 \rangle)| < \varepsilon/2$.

Now, let the sets V, W satisfy $P(V) = P(W) < \delta$. Put $V' = V \setminus W$, $W' = W \setminus V$. We can assume that $\delta < 1/6$; then we can find sets V'' and W'' such that V', W', V'', W'' are disjoint and which satisfy the following equalities

$$\begin{aligned} P(V' \cap K_1) &= P(V'' \cap K_2), & P(V' \cap K_2) &= P(V'' \cap K_1), \\ P(W' \cap K_1) &= P(W'' \cap K_2), & P(W' \cap K_2) &= P(W'' \cap K_1). \end{aligned}$$

Set

$$\begin{aligned} L_1 &:= (K_1 \cup V' \cup V'') \setminus (W' \cup W''), & L_2 &:= \Omega \setminus L_1, \\ L'_1 &:= L_1 \Delta V' \Delta W', & L'_2 &:= \Omega \setminus L'_1. \end{aligned}$$

Then

$$\begin{aligned} \mathcal{T}_{V'W'} \langle L_1, L_2 \rangle &= \langle L'_1, L'_2 \rangle, & P(L_1) &= P(L'_1) = 1/3, \\ P(K_1 \Delta L_1) &= 2P(V') < 2\delta, & P(K_1 \Delta L'_1) &\leq 4P(V') < 4\delta. \end{aligned}$$

From this we get

$$|\Delta(V, W)| = |\Delta(V', W', 2)| = |I(\langle L'_1, L'_2 \rangle) - I(\langle L_1, L_2 \rangle)| < \varepsilon.$$

□

Recall that P is defined on the σ -algebra \mathcal{F} . Let \mathcal{R} be the family of all sets in \mathcal{F} whose measure is rational.

Lemma 6. *Let $\Delta(V, W) \in \mathbb{R}$ be defined for any pair of sets $V, W \in \mathcal{R}$ of the same measure P and let it satisfy the following conditions*

(1) *For any sets U, V and W of the same measure*

$$\Delta(U, W) = \Delta(U, V) + \Delta(V, W).$$

(2) *For $V = V_1 + V_2$, $W = W_1 + W_2$, $P(V_i) = P(W_i)$, $i = 1, 2$ we have*

$$\Delta(V, W) = \Delta(V_1, W_1) + \Delta(V_2, W_2).$$

There is a unique finitely-additive set function $m: \mathcal{R} \rightarrow \mathbb{R}$ with $m(\Omega) = 0$ such that for any sets V and W with $V \cap W \in \mathcal{R}$ (i.e. belonging to the same algebra $\mathcal{G} \subset \mathcal{R}$) and satisfying $P(V) = P(W)$ we have

$$(9) \quad \Delta(V, W) = m(W) - m(V).$$

What is more, for any $A \in \mathcal{R}$ we have

$$(10) \quad |\mathbf{m}(A)| \leq \sup |\Delta(\cdot, A)|.$$

Proof. Let $A_1 + \dots + A_n = \Omega$ with $P(A_i) = 1/n$. Write

$$(11) \quad \mathbf{m}(A_1 + \dots + A_k) := \frac{1}{\binom{n}{k}} \sum_{\substack{\{i_1, \dots, i_k\} \\ \subset \{1, \dots, n\}}} \Delta(A_{i_1} + \dots + A_{i_k}, A_1 + \dots + A_k),$$

(where by writing $\{i_1, \dots, i_k\}$ we assume that these numbers are all distinct). Using property 2 of Proposition 5 we obtain

$$(12) \quad \mathbf{m}(A_1 + \dots + A_k) = \frac{1}{k! \binom{n}{k}} \sum_{\substack{(i_1, \dots, i_k): \\ \{i_1, \dots, i_k\} \subset \{1, \dots, n\} \\ 1 \leq j \leq k}} \Delta(A_{i_j}, A_j) = \frac{1}{n} \sum_{\substack{1 \leq i \leq n \\ 1 \leq j \leq k}} \Delta(A_i, A_j).$$

We show first that \mathbf{m} is well defined. Indeed, if $\mathcal{D} = \langle D_1, \dots, D_{ns} \rangle$, $P(D_p) = 1/(ns)$ and $\mathcal{A} = \langle A_1, \dots, A_n \rangle \subset \mathcal{D}$ say $A_i = D_{(i-1)s+1} + \dots + D_{is}$, $1 \leq i \leq n$ then

$$\sum_{\substack{(i-1)s+1 \leq p \leq is \\ (j-1)s+1 \leq q \leq js}} \Delta(D_p, D_q) = s \Delta(A_i, A_j).$$

Hence

$$(13) \quad \frac{1}{n} \sum_{\substack{1 \leq i \leq n \\ 1 \leq j \leq k}} \Delta(A_i, A_j) = \frac{1}{ns} \sum_{\substack{1 \leq p \leq ns \\ 1 \leq q \leq ks}} \Delta(D_p, D_q).$$

If we have now any other partition $B_1 + \dots + B_m = \Omega$, $P(B_j) = 1/m$ such that $A_1 + \dots + A_k = B_1 + \dots + B_l$ then there is a partition $C_1 + \dots + C_s$, $P(C_v) = 1/s$ such that $C_1 + \dots + C_r = A_1 + \dots + A_k$ and

$$\begin{aligned} \langle C_1, \dots, C_r \rangle &\perp_{A_1 + \dots + A_k} \langle A_1, \dots, A_k \rangle, \langle B_1, \dots, B_l \rangle \\ \langle C_{r+1}, \dots, C_s \rangle &\perp_{A_{k+1} + \dots + A_n} \langle A_{k+1}, \dots, A_n \rangle, \langle B_{l+1}, \dots, B_m \rangle. \end{aligned}$$

(Naturally, the symbol \perp_K denotes the conditional independence of algebras in $\mathfrak{A}|_K$, with respect to the truncated conditional measure $P|_K = P(\cdot)/P(K)$.) By the rationality of measure $P(D)$, where D is any atom of the algebra $\mathcal{A} \cdot \mathcal{C}$, we can find an algebra $\mathcal{D} \supset \mathcal{A}, \mathcal{C}$, all atoms of which D_1, \dots, D_d are of the same measure. In particular, d is a multiple of both n and s ; moreover we can assume that for some e we have $D_1 + \dots + D_e = A_1 + \dots + A_k$. Observe now that formula (13) gives

$$\frac{1}{n} \sum_{\substack{1 \leq i \leq n \\ 1 \leq j \leq k}} \Delta(A_i, A_j) = \frac{1}{d} \sum_{\substack{1 \leq v \leq d \\ 1 \leq w \leq e}} \Delta(D_v, D_w) = \frac{1}{s} \sum_{\substack{1 \leq h \leq s \\ 1 \leq k \leq r}} \Delta(C_h, C_k).$$

Similarly

$$\frac{1}{m} \sum_{\substack{1 \leq v \leq m \\ 1 \leq w \leq l}} \Delta(B_v, B_w) = \frac{1}{s} \sum_{\substack{1 \leq h \leq s \\ 1 \leq k \leq r}} \Delta(C_h, C_k).$$

This proves that \mathbf{m} is well defined.

The additivity of the mapping \mathbf{m} follows from equality (12). What is more we have $\Delta(V, W) = \mathbf{m}(W) - \mathbf{m}(V)$ for $V \cap W \in \mathcal{R}$. Indeed, if $V = A_1 + \dots + A_k$ and $W = A_{\sigma(1)} + \dots + A_{\sigma(k)}$ where $P(A_1) = \dots = P(A_n)$ and $\sigma(i) \in \{1, \dots, n\}$ then

$$\begin{aligned} \Delta(V, W) &= \frac{1}{\binom{n}{k}} \sum_{\substack{\{i_1, \dots, i_k\} \\ \subset \{1, \dots, n\}}} \Delta(A_1 + \dots + A_k, A_{\sigma(1)} + \dots + A_{\sigma(k)}) \\ &= \frac{1}{\binom{n}{k}} \sum_{\substack{\{i_1, \dots, i_k\} \\ \subset \{1, \dots, n\}}} \Delta(A_{i_1} + \dots + A_{i_k}, A_{\sigma(1)} + \dots + A_{\sigma(k)}) - \\ &\quad \Delta(A_{i_1} + \dots + A_{i_k}, A_1 + \dots + A_k) \\ &= \mathbf{m}(W) - \mathbf{m}(V). \end{aligned}$$

Definition (11) now gives $\mathbf{m}(\Omega) = 0$. Conversely, if \mathbf{m} satisfies $\mathbf{m}(\Omega) = 0$ and (9) it must satisfy (11) and so \mathbf{m} is unique. Inequality (10) follows from definition (11). \square

Lemma 7. *Let $\Delta(V, W) \in \mathbb{R}$ be defined for any pair of sets $V, W \in \mathcal{F}$ of the same measure P and let it satisfy conditions 1–3 of Proposition 5. There is a unique finitely-additive set function $\mathbf{m}: \mathcal{F} \rightarrow \mathbb{R}$, absolutely continuous with respect to measure P such that $\mathbf{m}(\Omega) = 0$ and such that for any sets V and W of the same measure P we have an equality*

$$\Delta(V, W) = \mathbf{m}(W) - \mathbf{m}(V).$$

What is more, for any $A \in \mathcal{F}$ we have $|\mathbf{m}(A)| \leq \sup|\Delta(\cdot, A)|$.

Proof. We will extend \mathbf{m} from Lemma 6 onto the whole algebra \mathcal{F} .

With the help of uniform continuity of Δ it is easy to show that for any measurable set A the supremum of $|\Delta(\cdot, A)|$ is finite and that the mapping $A \mapsto \sup|\Delta(\cdot, A)|$ is continuous.

Let now $\mathcal{R} \ni A_i \longrightarrow A \in \mathcal{F}$, i.e. $P(A \triangle A_i) \longrightarrow 0$. Then we have $\lim_{i,j \rightarrow +\infty} P(A_i \setminus A_j) = 0$, and then by (10) $\lim_{i,j \rightarrow +\infty} \mathbf{m}(A_i \setminus A_j) = 0$. Hence $\lim_{i,j \rightarrow +\infty} |\mathbf{m}(A_i) - \mathbf{m}(A_j)| = 0$, which means that the sequence $\mathbf{m}(A_i)$ converges. Put $\mathbf{m}(A) := \lim \mathbf{m}(A_i)$. Clearly $\mathbf{m}(A)$ is well defined; (indeed, by taking any other sequence $B_i \longrightarrow A$, it suffices to consider the combined sequence $A_1, B_1, A_2, B_2, \dots \longrightarrow A$ thereby getting the convergence of $\mathbf{m}(A_1), \mathbf{m}(B_1), \mathbf{m}(A_2), \dots$). We now get inequality (10) for any $A \in \mathcal{F}$; in particular it signifies the absolute continuity of \mathbf{m} with respect to P . The additivity of \mathbf{m} , the equality $\Delta(V, W) = \mathbf{m}(W) - \mathbf{m}(V)$ and the uniqueness of \mathbf{m} follow from analogical properties obtained previously for rational values of P . \square

Notice that the converse of the above lemma is trivially valid — having \mathbf{m} as in the thesis of Lemma, Δ satisfies conditions 1–3 of Proposition 5.

We shall go over now to the crucial theorem of this paper.

Theorem 1. *Every additive partition entropy I , continuous on \mathfrak{A}_2 has a unique decomposition into additive partition entropies:*

$$I = \tilde{I} + L_m,$$

where \tilde{I} is an additive partition entropy depending solely on the measures of atoms, whereas L_m is a continuous additive partition entropy from Example 1 for some finitely-additive set function m absolutely continuous with respect to P such that $m(\Omega) = 0$. If I is continuous, then so is \tilde{I} .

Proof. Let m will be as in Lemma 7 for Δ of Proposition 5. Observe at first that the additive partition entropy \tilde{I} defined by

$$\tilde{I}(\mathcal{A}) := I(\mathcal{A}) - L_m(\mathcal{A}) = I(\mathcal{A}) - \left[m(A_1) \log \frac{1}{P(A_1)} + \cdots + m(A_n) \log \frac{1}{P(A_n)} \right],$$

where $\mathcal{A} = \langle A_1, \dots, A_n \rangle$, is invariant in the following sense

$$\tilde{I}(\mathcal{T}_{VW}\mathcal{A}) = \tilde{I}(\mathcal{A}) \quad \text{for } V, W \text{ such that } \mathcal{A} \in \mathfrak{F}_{VW}, P(V) = P(W).$$

Indeed, after assuming that $V \subset A_1$, $W \subset A_2$, setting

$$B_i := A_i \Delta V \Delta W \in \mathcal{T}_{VW}\mathcal{A}, \quad i = 1, 2$$

and using the equalities

$$\Delta(V, W) = m(W) - m(V) = m(B_1) - m(A_1) = m(A_2) - m(B_2)$$

we obtain

$$\begin{aligned} I(\mathcal{T}_{VW}\mathcal{A}) - I(\mathcal{A}) &= \Delta(V, W) \cdot \log \frac{P(A_2)}{P(A_1)} \\ &= (m(B_1) - m(A_1)) \log \frac{1}{P(A_1)} - (m(A_2) - m(B_2)) \log \frac{1}{P(A_2)} \\ &= \left[\begin{array}{c} m(B_1) \log \frac{1}{P(B_1)} \\ + m(B_2) \log \frac{1}{P(B_2)} \end{array} \right] - \left[\begin{array}{c} m(A_1) \log \frac{1}{P(A_1)} \\ + m(A_2) \log \frac{1}{P(A_2)} \end{array} \right] \\ &= L_m(\mathcal{T}_{VW}\mathcal{A}) - L_m(\mathcal{A}). \end{aligned}$$

It remains to show that by a sequence of operations of shape \mathcal{T}_{VW} , like above, we can go from any algebra \mathcal{A} to any other algebra \mathcal{B} , with the same measures of atoms as \mathcal{A} . Let $\mathcal{A} = \langle A_1, \dots, A_n \rangle$ and $\mathcal{B} = \langle B_1, \dots, B_n \rangle$, $P(A_i) = P(B_i)$. According to Corollary 1 (see Remark 2) we may assume that $\mathcal{A} \perp \mathcal{B}$. Then the sets $K_{i,j} := A_i \cap B_j$ are disjoint and $P(K_{i,j}) = P(K_{j,i})$; moreover for $i \neq j$ the sets $K_{i,j}$ and $K_{j,i}$ are contained in distinct atoms A_i , A_j of the partition $\{A_1, \dots, A_n\}$. It is easy to see that we have the following equality:

$$\mathcal{B} = \mathcal{T}_{K_{n-1,n}K_{n,n-1}} \cdots \mathcal{T}_{K_{1,3}K_{3,1}} \mathcal{T}_{K_{1,2}K_{2,1}} \mathcal{A}.$$

The uniqueness of the decomposition follows from Remark 1 on page 7. The continuity of the additive partition entropy L_m is described in Remark 6. \square

5. FINAL REMARKS

At first, let us mention some other possible approaches one might take to derive our results. The first prototype of the proof of Lemma 7 was based around defining $\widehat{\mathbf{m}}(1_V - 1_W) := \Delta(V, W)$, extending $\widehat{\mathbf{m}}$ to the linear space generated by functions of shape 1_V , observing that $\widehat{\mathbf{m}}$ is a continuous linear mapping, and then extending $\widehat{\mathbf{m}}$ to the space of all continuous functions and finally defining $\mathbf{m}(V) := \widehat{\mathbf{m}}(1_V)$.

One could try to prove Theorems 1 by elaborating on Remark 3 and a version of Proposition 3. Although unexplored, it seems that such an approach would suffer from some problems of its own.

We would like now to convert our theorem to the typical case of a countably-additive probability. To do so, we need the following simple and known properties of measure. The proofs are omitted.

Remark 8. *If \mathbf{P} is a countably-additive measure on σ -algebra \mathcal{F} then every finitely-additive set function \mathbf{m} absolutely continuous with respect to \mathbf{P} is countably-additive.*

Remark 9. *Every nonatomic probability space satisfies the Darboux condition.*

In view of these remarks, we can recast Theorem 1 in the setting of a nonatomic probability space $(\Omega, \Sigma, \mathbf{P})$.

Theorem 2. *Let I be a additive partition entropy, continuous on \mathfrak{A}_2 (c.f. Definition 3, and Definition 5). There exists an additive entropy, and a countably-additive set function \mathbf{m} absolutely continuous with respect to probability \mathbf{P} such that*

$$I = H_{\mathbf{P}} + L_{\mathbf{m}},$$

There is only one such pair with $\mathbf{m}(\Omega) = 0$. $L_{\mathbf{m}}$ is continuous. If I is continuous then also $H_{\mathbf{P}}$ is continuous, i.e. H is continuous in the open domain. (c.f. Remark 7)

We would like to mention that this result will be transferred to a quantum context in a follow-up paper, [20].

REFERENCES

- [1] ACZÉL, J. (1978). A mixed theory of information-II: Additive inset entropies (of randomized systems of events) with measurable sum property. *Utilitas Math.* **13** 49–54.
- [2] ACZÉL, J. and DARÓCZY, Z. (1975). *On measures of information and their characterizations*. Academic Press, Inc.
- [3] ACZÉL, J. and DARÓCZY, Z. (1978). A mixed theory of information-I: Symmetric, recursive and measurable entropies of randomized system of events. *RAIRO Informat. Theor.* **12** 149–155.
- [4] ACZÉL, J., FORTE, B., and NG, C. T. (1974). Why the Shannon and Hartley entropies are 'natural'. *Adv. in Appl. Probab.* **6** 131–146.

- [5] ARTSTEIN, P. (1992). Sensitivity to σ -fields of information in stochastic allocation. *Stochastics and Stochastics Reports* **36** 41–63.
- [6] BAEZ, J. C., FRITZ, T., and LEINSTER, T. A characterization of entropy in terms of information loss. (*preprint*)
- [7] BEHARA M. , and NATH, P. (1973). Additive and Non-additive entropies of finite measurable partitions. in *Probablity and Information theory II*, *LNM* **296** 102–138.
- [8] BENNET, C., BRASSARD, G., CRÉPEAU, C., and MAURER, U. (1995). Generalized privacy amplification. *IEEE Trans. Inf. Theory* **41** 1915–1923
- [9] CAMPBELL, L. L. (1965). A coding theorem and Rényi’s entropy. *Information and Control* **8** 423–429
- [10] CAMPBELL, L. L. (1985). The relation between information theory and the differential geometry approach to statistics *Inform. Sci.* **35** 199–210.
- [11] COVER, T. M. and THOMAS, J. A. (1991). *Elements of Information Theory*. John Wiley & Sons, Inc.
- [12] CSISZÁR, I. (2008). Axiomatic Characterizations of Information Measures. *Entropy* **10** 261–273.
- [13] EBANKS, B. R. (1986). Measures of inset information on open domains-II: additive inset entropies with measurable sum property. *Prob. Th. Rel. Fields* **73** 517–528.
- [14] FADEEV, D. K. (1956). Towards the notion of finite entropy of a system of probabilities. (in russian) *Uspiehi. matem. nauk.* **11** 227–231.
- [15] GUIAŞU, S. and THEODORESCU, R. (1986). *La Théorie Mathématique de l’Information*. Dunod, Paris.
- [16] HARTLEY, R. V. (1928). Transmission of information *Bell System Tech. J.* **7** 535–563.
- [17] KANNAPPAN, P. and SANDER, W. (2004). Inset information measures on open domains. *Aequationes Math.* **68** 289–320.
- [18] NG, C. T., LUCE, R. D., and MARLEY, A. (2007). Utility of Gambling when Events are Valued: an Application of Inset Entropy. *Theory and Decision* **67** 23–63.
- [19] PASZKIEWICZ, A. *On generaliztion and pricing of information*. (in preparation)
- [20] PASZKIEWICZ, A., and SOBIESZEK, T. *On quantum information*. (*preprint*)
- [21] RÉNYI, A. (1960). On measures of entropy and information *Proc. 4th Berkeley Symp. Math. Stat. Probab.* **1** 547–561.
- [22] RÉNYI, A. (1965). On the foundations of information theory *Rev. Inst. Internat. Stat.* **33**, 1–4.
- [23] SHANNON, C. E., and WEAVER, W (1964). *The mathematical theory of communication*. The University of Illinois press.
- [24] SIMOVICI, D. A., and JAROSZEWICZ, S. (2002). An axiomatization of partition entropy. *IEEE Trans. Inf. Theory*, **48** 2138–2142.

- [25] SOBIESZEK, T. *Noncontinous additive entropies of partitions* (preprint, [arXiv:1202.4590 \[cs.IT\]](#))
- [26] TRAUB J. F., WOŹNIAKOWSKI, H., and WASILKOWSKI, G. W. (1988). *Information-based complexity*. Academic Press, Inc.

FACULTY OF MATHEMATICS AND COMPUTER SCIENCE, UNIVERSITY OF ŁÓDŹ, UL.
BANACHA 22, 90-238 ŁÓDŹ, POLAND

E-mail address, A. Paszkiewicz: ktpis@math.uni.lodz.pl

E-mail address, T. Sobieszek: sobieszek@math.uni.lodz.pl

URL: <http://sobieszek.co.cc>